

Fitting Markovian binary trees using global and individual demographic data

Sophie Hautphenne*, Melanie Massaro[†] and Katharine Turner [‡]

February 15, 2017

Abstract

We consider a class of branching processes called Markovian binary trees, in which the individuals lifetime and reproduction epochs are modeled using a transient Markovian arrival process (TMAP). We estimate the parameters of the TMAP based on population data containing information on age-specific fertility and mortality rates. Depending on the degree of detail of the available data, a weighted non-linear regression method or a maximum likelihood method is applied. We discuss the optimal choice of the number of phases in the TMAP, and we provide confidence intervals for the model outputs. The results are illustrated using real data on human and bird populations.

Keywords: Markovian binary tree; transient Markovian arrival process; Markov modulated Poisson process; parameter estimation; non-linear regression; maximum likelihood; petroica traversi

1 Introduction

Simple birth-and-death processes do not offer enough flexibility to model real biological populations in which the age of individuals impacts on their fertility and mortality rates. The memoryless property inherent to these models implies that individuals do not age. However, they are tractable

*The University of Melbourne and Ecole polytechnique fédérale de Lausanne, sophiemh@unimelb.edu.au

[†]Charles Sturt University, mmassaro@csu.edu.au

[‡]Ecole polytechnique fédérale de Lausanne, katharine.turner@epfl.ch

and amenable to efficient parameter estimation. In this paper, we model the lifetime and reproduction epochs of individuals in a population using a *transient Markovian arrival process* (TMAP). Roughly speaking, a TMAP is a point process in which the event rate depends on the state of an underlying transient Markov chain with n transient states (also called *phases*), and one absorbing phase. Each event in the TMAP corresponds to the birth of a child, and the absorption in phase 0 corresponds to the individual's death. The resulting continuous-time branching process, called *Markovian binary tree* (MBT), is the matrix generalisation of the birth-and-death process. It allows for much more flexibility than the latter, while keeping an excellent computational tractability.

Performance measures of MBTs include the extinction probability of the population, the distributions of the time until extinction, the population size at any given time, and the total progeny size until any given time. The MBT model has already been used to efficiently compare demographic properties of female families in different countries, see [4]. The motivation behind the present paper is to develop the statistical tools necessary to fit an MBT to populations of species for which detailed information about age-specific survival and reproductive rates of individuals is available. The model can then be used to calculate age-dependent demographic properties. By knowing the exact age of individuals of a population, its future survival probability can be assessed, which aids conservation management of endangered species.

We fit a TMAP to different types of datasets which may be available from demographic databases or from studying an animal population in the field. These datasets can have different degrees of detail. We distinguish between:

- *Global population data*, consisting of the *average* age-specific fertility and mortality *rates* over an entire population. This sort of data is usually provided in databases on human fertility and mortality.
- *Individual demographic data*, consisting of data on age-specific fertility and mortality *counts* for each individual in a population. This sort of data often exists for closely monitored animal species. Here we will use data from a highly threatened bird species, the Chatham Island black robin (*Petroica traversi*) [1, 10].

The parameter estimation method depends on the type of data which are available: in the global population case, we use a weighted non-linear regression method to fit the parameters, and in the individual demographic data

case, we use a maximum likelihood method. We consider different validation methods to determine the optimal number of phases n in the TMAP. Once a value of n is determined and an estimator is found for the model parameters, we derive confidence intervals for the model outputs.

We apply our results to two real-world examples. The weighted non-linear regression method is applied in human demography leading to an improvement of the Markovian model considered in [4]. The maximum likelihood method is applied to the black robin population providing important insights about the species demography to be further discussed from a conservation biology point of view in a subsequent paper.

The paper is organised as follows: in Section 2, we introduce TMAPs and describe the special case that we shall focus on. In Section 3 we perform model parameter estimation based on the average age-specific fertility and mortality rates, and in Section 4, we estimate the parameters based on individual age-specific fertility and mortality counts. In Section 5 we apply each method on a real-world example.

2 Transient Markovian arrival processes

Transient Markovian arrival processes (TMAPs) are two-dimensional Markovian processes $\{(N(t), \varphi(t)) : t \in \mathbb{R}^+\}$ on the state space $\mathbb{N} \times \{0, 1, \dots, n\}$, where n is finite, combining the *level* process $\{N(t)\}$, which counts the number of arrivals in $[0, t]$, with the *phase* process, $\{\varphi(t)\}$, which is a continuous-time Markov chain. The states $(k, 0)$ are absorbing for all $k \geq 0$; the other states are transient.

A TMAP is characterized by two $n \times n$ rate matrices D_0 and D_1 and a non-negative $n \times 1$ rate vector \mathbf{d} . Feasible transitions are from (k, i) to (k, j) , for $k \geq 0$ and $1 \leq i \neq j \leq n$ at the rate $(D_0)_{ij}$, or from (k, i) to $(k+1, j)$ for $1 \leq i, j \leq n$ at the rate $(D_1)_{ij}$, or from (k, i) to $(k, 0)$ at rate d_i . The first transitions (at rate $(D_0)_{ij}$) are *hidden*: the phase of the individual changes but the level is not incremented. The second transitions (at rate $(D_1)_{ij}$) are *observable*: a birth (arrival) is recorded, and the state of the individual may or may not change. The third transitions (at rate d_i) indicate the termination of the individual's life.

The matrix D_1 and the vector \mathbf{d} are nonnegative, D_0 has nonnegative off-diagonal elements and strictly negative elements on the diagonal such that $D_0 \mathbf{1} + D_1 \mathbf{1} + \mathbf{d} = \mathbf{0}$, where $\mathbf{1}$ is an $n \times 1$ vector of ones. One also defines

the initial probability vector $\boldsymbol{\alpha} = (\alpha_i)_{1 \leq i \leq n}$, and we assume that $\boldsymbol{\alpha} \mathbf{1} = 1$, so that $\varphi(0) \neq 0$ a.s. More details on TMAPs can be found in [8].

There is a total of $p = 2n^2 + n - 1$ entries in the matrices $\boldsymbol{\alpha}$, D_0 , D_1 , \mathbf{d} if no assumption is made on their structure. A special case of TMAP, called the *acyclic transient Markov modulated Poisson process* (ATMMPP), assumes

- individuals start their lifetime in phase 1 with probability one,
- they can only move from phase i to phase $i + 1$ or to phase 0, with respective rates γ_i for $1 \leq i \leq n - 1$ and d_i for $1 \leq i \leq n$,
- while in phase i , they reproduce at rate λ_i and do not make any simultaneous phase transition at reproduction time.

With these assumptions we have $\boldsymbol{\alpha} = [1, 0, \dots, 0]$, $D_1 = \text{diag}(\lambda_1, \dots, \lambda_n)$, and the only non-zero entries of D_0 are $(D_0)_{i,i+1} = \gamma_i$ and

$$(D_0)_{ii} = \begin{cases} \lambda_i - d_i - \gamma_i, & 1 \leq i \leq n - 1, \\ -\lambda_i - d_i, & i = n. \end{cases}$$

There is a total of $p = 3n - 1$ parameters in an ATMMPP.

The lifetime distribution of a TMAP is phase-type $\text{PH}(\boldsymbol{\alpha}, D_0 + D_1)$, see [8]. Due to the structure of D_0 and D_1 in the ATMMPP case, this corresponds to a Coxian distribution. Such distributions are important as any acyclic phase-type distribution has an equivalent Coxian representation. Therefore, in terms of the lifetime distribution, the ATMMPP does not impose much restriction compared to the general TMAP.

3 Global population data

3.1 Available data and model equivalent

For this section, we assume the available data are estimates of the expected age-specific fertility rates, \hat{b}_x , and estimates of the expected age-specific mortality rates, \hat{d}_x , where $x \in \{0, 1, 2, \dots, M\}$ denotes the age, that is, the period of time $[x, x + 1)$ during the lifetime, and M is the maximal age for which data are available. The method developed in this section can be generalised to ℓ -year age classes, details are provided in Section 7.2 of the Supplementary Material.

The rates \hat{b}_x and \hat{d}_x are interpreted respectively as the expected number of offspring per year from a parent at age x and the probability that an individual who reached age x dies within the year. We denote by $\bar{d}(x)$ and $\bar{b}(x)$ the equivalent quantities computed from the TMAP model. These functions have the following analytic expression, the proof of which is provided in Section 7.1 of the Supplementary Material.

Proposition 3.1. *The age-specific mortality and fertility rates in a TMAP with phase transition rate matrix $D := D_0 + D_1$ are respectively given by*

$$\begin{aligned}\bar{d}(x) &= \frac{\alpha e^{Dx}(I - e^D)\mathbf{1}}{\alpha e^{Dx}\mathbf{1}} \\ \bar{b}(x) &= \frac{\alpha e^{Dx}(I - e^D)(-D)^{-1}D_1\mathbf{1}}{\alpha e^{Dx}\mathbf{1}}.\end{aligned}$$

3.2 Parameter estimation

In [9] only death rates were used to fit phase-type lifetime distributions. We extend this approach, estimating the model parameters by minimizing the sum of weighted squared errors

$$F = \sum_{x=0}^M \left[(\hat{d}_x - \bar{d}(x))^2 + (\hat{b}_x - \bar{b}(x))^2 \right] \hat{S}_x, \quad (3.1)$$

where the weights \hat{S}_x are the observed probabilities of survival until age x ,

$$\hat{S}_x = (1 - \hat{d}_0)(1 - \hat{d}_1) \cdots (1 - \hat{d}_{x-1}).$$

As age increases there may be less available data leading to higher variance. These weights balance the potential resulting heteroscedasticity. If the estimated age-specific rates \hat{d}_x and \hat{b}_x are computed as averages of n_x uncorrelated raw observations, another simple choice of weights would be $W_x = n_x$.

Since the functions $\bar{d}(x)$ and $\bar{b}(x)$ are non-linear in both the input variable x and in the parameters of the TMAP, we are dealing with a weighted non-linear regression. If there is missing information in the data and no estimate exists for \hat{d}_x or \hat{b}_x for some age x , then we set the corresponding term $(\hat{d}_x - \bar{d}(x))^2$ or $(\hat{b}_x - \bar{b}(x))^2$ to zero in the sum.

Remark 3.1. The function $\bar{d}(x)$ corresponds to the hazard rate at age x in survival analysis. Several hazard models have been considered to fit mortality

data, such as the Gompertz-Makeham, the Siler, and the Heligman-Pollard models [3]. Similarly, several age-specific fertility models have been studied, including the Hadwiger, the Beta, and the Gamma models [13]. The functions $\bar{d}(x)$ and $\bar{b}(x)$ are not claimed to provide better fits than these models, however, as opposed to the known mortality and fertility models which are generally studied separately, $\bar{d}(x)$ and $\bar{b}(x)$ are performance measures coming from *the same Markovian model*, and are thus optimised together. The estimated Markovian model then corresponds to the best model fitting both the mortality and fertility data, and can be used to make a complete demographic study of the population, as shown in [4].

3.3 Goodness of fit and optimal value of n

In the global population data case we estimate the model parameters by minimizing the objective function (3.1). Therefore, a natural choice for the *mean square error* function is

$$\text{MSE} = E \left[\sum_{x=0}^M \left[(\bar{d}(x) - \hat{\bar{d}}(x))^2 + (\bar{b}(x) - \hat{\bar{b}}(x))^2 \right] \bar{S}(x) \right].$$

Here $\bar{d}(x)$, $\bar{b}(x)$, and $\bar{S}(x)$ are respectively the age-specific mortality function, the age-specific fertility function, and the age-specific survival function corresponding to the *true* model, and $\hat{\bar{d}}(x)$ and $\hat{\bar{b}}(x)$ are the equivalent functions corresponding to the *estimated* model. If we know the true model then the MSE can be estimated through resampling. Alternatively this could be estimated when we are given a collection of datasets each containing global population information.

The value of the MSE indicates of the goodness of fit of the model. When the true model is unknown we estimate the optimal number of phases n by minimizing the MSE. When the true model is known, comparing the MSE for different values of n informs us on the sensitivity of the output with respect to the number of phases, as illustrated in Section 7.4 in the Supplementary Material.

4 Individual demographic data

4.1 Available data

In this section, we assume the data are individual age-specific fertility and mortality counts in successive age-classes of length $\ell > 0$ ¹. They consist of N vectors (one for each individual) of the type

$$\mathbf{v} = [6, 8, -2, 9, 0, 3, 3, -1], \quad (4.1)$$

of variable length, whose entries $v_i, i \geq 1$ are interpreted as follows:

- $v_i = k \in \{0, 1, 2, \dots\}$ if the individual had k offspring while in the age-class $[(i-1)\ell, i\ell)$,
- $v_i = -1$ if the individual died in the previous age-class $[(i-2)\ell, (i-1)\ell)$, possibly after producing some offspring, and
- $v_i = -2$ if the individual was alive at the beginning of the age-class $[(i-1)\ell, i\ell)$ but there is no (or incomplete) information on her progeny in that age-class.

4.2 Parameter estimation

Based on a sample of i.i.d. individual life vectors $\{\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(N)}\}$, we maximize the log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(N)}) = \sum_{j=1}^N \log p(\mathbf{v}^{(j)} | \boldsymbol{\theta}), \quad (4.2)$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, D_0, D_1, \mathbf{d}\}$, and $p(\mathbf{v}^{(j)} | \boldsymbol{\theta})$ is the probability of observing the individual life vector $\mathbf{v}^{(j)}$, under the model parameter $\boldsymbol{\theta}$.

Let $K = \max_{i,j} \{v_i^{(j)} : 1 \leq i, 1 \leq j \leq N\}$ be the maximum number of offspring per age-class among the individuals in the sample. The probabilities $p(\mathbf{v}^{(j)} | \boldsymbol{\theta})$ can be written as matrix products involving the matrices and vectors

¹Successive age-classes are assumed of equal length, but though computationally convenient, this is not essential.

$P(k) = (P_{ij}(k))$, $\mathbf{p}(k) = (p_i(k))$, $P = (P_{ij})$, and $\mathbf{p} = (p_i)$ defined as

$$P_{ij}(k) := P[N(\ell) = k, \varphi(\ell) = j | N(0) = 0, \varphi(0) = i], \quad (4.3)$$

$$p_i(k) := P[N(\ell) = k, \varphi(\ell) = 0 | N(0) = 0, \varphi(0) = i], \quad (4.4)$$

$$P_{ij} := P[\varphi(\ell) = j | \varphi(0) = i] = \sum_{k \geq 0} P_{ij}(k), \quad (4.5)$$

$$p_i := P[\varphi(\ell) = 0 | \varphi(0) = i] = \sum_{k \geq 0} p_i(k), \quad (4.6)$$

for $1 \leq i, j \leq n$ and $1 \leq k \leq K$. As an illustrative example, consider the four life vectors

$$\mathbf{v}^{(1)} = [2, 3, 1, -1], \quad \mathbf{v}^{(2)} = [2, -2, 1, -1], \quad \mathbf{v}^{(3)} = [2, 3], \quad \mathbf{v}^{(4)} = [2, -2].$$

By conditioning on the phases of the TMAP at the boundaries of the successive ℓ -year intervals, the probability of observing these vectors is

$$\begin{aligned} p(\mathbf{v}^{(1)} | \boldsymbol{\theta}) &= \boldsymbol{\alpha} P(2) P(3) \mathbf{p}(1), & p(\mathbf{v}^{(2)} | \boldsymbol{\theta}) &= \boldsymbol{\alpha} P(2) P \mathbf{p}(1), \\ p(\mathbf{v}^{(3)} | \boldsymbol{\theta}) &= \boldsymbol{\alpha} P(2) [P(3) \mathbf{1} + \mathbf{p}(3)], & p(\mathbf{v}^{(4)} | \boldsymbol{\theta}) &= \boldsymbol{\alpha} P(2) \mathbf{1}. \end{aligned}$$

Note that if \mathbf{v} is a vector of size $M + 1$ with all entries equal to -2 , then $p(\mathbf{v} | \boldsymbol{\theta}) = \boldsymbol{\alpha} P^M \mathbf{1}$ is the probability that the individual survives at least the first M age-classes.

The quantities defined in (4.3)–(4.6) can be computed explicitly, as shown in the next proposition, whose proof is provided in Section 7.3 of the Supplementary Material.

Proposition 4.1. *For $1 \leq k \leq K$, the matrix $P(k)$ and the vector $\mathbf{p}(k)$ are given by*

$$\begin{aligned} P(k) &= (1/k!) (\mathbf{e}_k \otimes I) \exp(\mathcal{M}\ell) (\mathbf{e}_1^\top \otimes I), \\ \mathbf{p}(k) &= (1/k!) (\mathbf{e}_k \otimes I) [I - \exp(\mathcal{M}\ell)] (-\mathcal{M})^{-1} (\mathbf{e}_1^\top \otimes I) \mathbf{d}, \end{aligned}$$

where \mathbf{e}_k is the k th unit row vector of size K , and

$$\mathcal{M} = \begin{bmatrix} D_0 & & & & \\ D_1 & D_0 & & & \\ & 2D_1 & D_0 & & \\ & & \ddots & \ddots & \\ & & & KD_1 & D_0 \end{bmatrix}.$$

The matrix P and vector \mathbf{p} are given by

$$P = \exp(D\ell), \quad \mathbf{p} = [I - \exp(D\ell)] (-D)^{-1} \mathbf{d}.$$

Our MLE method generalises the results of Davison and Ramesh [2] who considered the parameter estimation of Markov modulated Poisson processes in the binary data case $v_i = \mathbb{1}_{\{N(i\ell) - N((i-1)\ell) \geq 1\}}$. To our knowledge, other parameter estimation methods for such processes are based on the observation of the successive inter-event times rather than on the number of events within successive time intervals, see for instance [14] and [15]. As confirmed in Figure 10, when the length of the time intervals decreases to zero, the estimates obtained with our method converge to those obtained with the usual method based on the observation of the successive inter-event times.

4.3 Goodness of fit and optimal number of phases

We consider three different criteria for choosing the optimal value of the number n of phases. These criteria are compared on numerical examples in Section 7.4 in the Supplementary Material.

Akaike Information Criterion (AIC)

We choose the value of n which minimizes the AIC defined as

$$\text{AIC} = 2p - 2\mathcal{L}(\hat{\boldsymbol{\theta}}; \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(N)}),$$

where the number of parameters $p = 3n - 1$ an ATMMPP model with n phases. This criterion deals with the trade-off between goodness of fit of the model and its complexity. One advantage is that it does not rely on the knowledge of the true model: the value of AIC can be directly computed from the log-likelihood of the estimated model given the data.

Cross-validation (CV)

We perform a K -fold cross-validation over the data sample of individual life vectors (with typical value $K = 5$). The idea is to randomly divide the data into K equal-sized parts. We leave out part k , fit the model to the other $K - 1$ parts (combined), and then evaluate the likelihood of the left-out k th part (test set) under the estimated parameters. We choose the model maximizing the mean test likelihood obtained by averaging the results for $k = 1, 2, \dots, K$. Similar to the AIC, this method does not require us to know the true model.

Mean squared integrated loss (MSIL)

Let \mathcal{V} be the set of all life vectors with entries in $\mathbb{N} \cup \{-1\}$. Any life vector with at least one entry equal to -2 is interpreted as a (disjoint) union of vectors in \mathcal{V} . For any fixed number of phases n , and a given sample of life vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$, the MLE method is used to estimate the parameters $\boldsymbol{\theta}_n = \{\boldsymbol{\alpha}, D_0, D_1, \mathbf{d}\}$ of the TMAP model. From the estimate $\hat{\boldsymbol{\theta}}_n$ we define a corresponding probability mass function $\hat{f}_n(\cdot)$ over \mathcal{V} as

$$\hat{f}_n(\mathbf{v}) = p(\mathbf{v}|\hat{\boldsymbol{\theta}}_n), \quad \mathbf{v} \in \mathcal{V}.$$

The optimal number of phases is the value of n minimizing the *mean squared integrated loss*, defined as

$$\begin{aligned} \text{MSIL} &= \mathbb{E} \left[\sum_{\mathbf{v} \in \mathcal{V}} (f(\mathbf{v}) - \hat{f}_n(\mathbf{v}))^2 \right] \\ &= \sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v})^2 - 2\mathbb{E} \left[\sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v}) \hat{f}_n(\mathbf{v}) \right] + \mathbb{E} \left[\sum_{\mathbf{v} \in \mathcal{V}} \hat{f}_n(\mathbf{v})^2 \right]. \end{aligned} \quad (4.7)$$

Since the first term is independent of n , the value of n minimizing $\text{MSIL}^*(n) := \mathbb{E}[\sum_{\mathbf{v} \in \mathcal{V}} \hat{f}_n(\mathbf{v})^2] - 2\mathbb{E}[\sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v}) \hat{f}_n(\mathbf{v})]$ also minimizes the MSIL. The problem therefore reduces to estimating $\text{MSIL}^*(n)$ for each n .

If the true model is known, then $f(\cdot)$ is known, and the expectations in $\text{MSIL}^*(n)$ can be estimated through resampling. If the true model is unknown, then a K -fold cross-validation method can be applied to estimate $\text{MSIL}^*(n)$. Let A_k and B_k be the k th training set and test set, respectively. Let $\hat{f}_n^k(\cdot)$ denote the probability mass function estimator using n phases and training set A_k . We have

$$\mathbb{E} \left[\sum_{\mathbf{v} \in \mathcal{V}} \hat{f}_n(\mathbf{v})^2 \right] \approx \frac{1}{K} \sum_{k=1}^K \sum_{\mathbf{v} \in \mathcal{V}} \hat{f}_n^k(\mathbf{v})^2, \quad (4.8)$$

and since the sets B_k are all drawn from the true distribution $f(\cdot)$, we have

$$\mathbb{E} \left[\sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v}) \hat{f}_n(\mathbf{v}) \right] \approx \frac{1}{K} \sum_{k=1}^K \frac{1}{|B_k|} \sum_{\mathbf{v} \in B_k} \hat{f}_n^k(\mathbf{v}).$$

The set of life vectors \mathcal{V} being infinite, the sums in (4.7) and (4.8) need to be modified in practice. For a given pair of integers K and M , we

partition the set \mathcal{V} to form a new *finite* set $\tilde{\mathcal{V}}_{K,M}$ such that $\sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v}) = \sum_{\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{K,M}} f(\tilde{\mathbf{v}}) = 1$. The vectors $\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{K,M}$ are of length M and have their entries in the finite set $\{-1, 0, 1, 2, \dots, K, K+1\}$, so that

$$|\tilde{\mathcal{V}}_{K,M}| = (K+2) \frac{(K+2)^{(M+1)} - 1}{(K+1)}.$$

They define equivalence classes in \mathcal{V} as follows:

- if $-1 \leq \tilde{v}_i \leq K$ for all $1 \leq i \leq M$, then

$$\tilde{\mathbf{v}} := \{\mathbf{v} \in \mathcal{V} : v_i = \tilde{v}_i, \text{ for all } i \in \{1, \dots, M\}\},$$

in which case $f(\tilde{\mathbf{v}}) = p(\tilde{\mathbf{v}}|\boldsymbol{\theta})$;

- if $\tilde{v}_{i_1} = \dots = \tilde{v}_{i_\ell} = K+1$ for some $1 \leq i_1, \dots, i_\ell \leq M$, $\ell \geq 1$, then

$$\begin{aligned} \tilde{\mathbf{v}} := \{ & \mathbf{v} \in \mathcal{V} : v_i = \tilde{v}_i \text{ for all } i \in \{1, \dots, M\} \setminus \{i_1, \dots, i_\ell\}, \\ & \text{and } v_i \geq K+1 \text{ for all } i \in \{i_1, \dots, i_\ell\}\}, \end{aligned}$$

in which case $f(\tilde{\mathbf{v}})$ is computed as given in the next Lemma.

Lemma 4.1. *For any $\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{K,M}$ such that $\tilde{v}_{i_1}, \dots, \tilde{v}_{i_\ell} = K+1$ for some indices $1 \leq i_1, \dots, i_\ell \leq M$, $\ell \geq 1$, we have*

$$f(\tilde{\mathbf{v}}) = \sum_{k_1, \dots, k_\ell \in \{-2, 0, 1, \dots, K\}} (-1)^{\ell + \sum_{i=1}^{\ell} \mathbb{1}_{\{k_i = -2\}}} p(\mathbf{v}^*(k_1, \dots, k_\ell)|\boldsymbol{\theta}), \quad (4.9)$$

where the vector $\mathbf{v}^*(k_1, \dots, k_\ell)$ is such that, for $1 \leq i \leq M$,

$$v_i^*(k_1, \dots, k_\ell) = \begin{cases} \tilde{v}_i & \text{if } i \notin \{i_1, \dots, i_\ell\} \\ k_j & \text{if } i = i_j \text{ for some } 1 \leq j \leq \ell. \end{cases}$$

Proof. We know $f(\tilde{\mathbf{v}}) = \sum_{k_1, \dots, k_\ell \geq K+1} p(\mathbf{v}^*(k_1, \dots, k_\ell)|\boldsymbol{\theta})$ by the definition of $\tilde{\mathbf{v}}$ and $\mathbf{v}^*(k_1, \dots, k_\ell)$. This sum contains ℓ embedded sums of the form $\sum_{k_j \geq K+1}$, which we rewrite as $\sum_{k_j \geq K+1} = \sum_{k_j \geq 0} - \sum_{0 \leq k_j \leq K}$. Using

$$\sum_{k_j \geq 0} p(\mathbf{v}^*(k_1, \dots, k_j, \dots, k_\ell)|\boldsymbol{\theta}) = p(\mathbf{v}^*(k_1, \dots, -2, \dots, k_\ell)|\boldsymbol{\theta}),$$

and rearranging the terms then lead to (4.9). \square

Replacing \mathcal{V} by $\tilde{\mathcal{V}}_{K,M}$ results in a different version of the MSIL criterion which selects the best model capturing differences in the number of children

less than or equal to K over the first M age-classes. It is clear that the larger the age-class length ℓ , the smaller M and the larger K should be chosen in order for \mathcal{V} to be well approximated by $\tilde{\mathcal{V}}_{K,M}$. When $\ell = 1$, a possible choice of the partitioning parameters is taking M as the ceiling of the expected lifetime plus one, and $K + 1$ as the maximal expected number of children per age-class, that is,

$$M = \lceil \sum_{x \geq 1} \hat{S}_x \rceil + 1, \quad \text{and} \quad K + 1 = \lceil \max_{1 \leq x \leq M} \hat{b}_x \rceil. \quad (4.10)$$

Another choice leading to smaller equivalence classes in \mathcal{V} is

$$M = \min\{x \geq 0 : \hat{S}_x < p\} + 1, \quad \text{and} \quad K + 1 = \lceil \max_{1 \leq x \leq M} (\hat{b}_x + \hat{\sigma}_x) \rceil, \quad (4.11)$$

where $1 - p$ is a covering probability, and $\hat{\sigma}_x$ is the standard error of the age-specific fertility rate at age x . Formulae for ℓ -year age-classes are analogous.

4.4 Confidence intervals for the model outputs

For any performance measure of the model $g(x, \boldsymbol{\theta})$, such as the mortality or fertility functions at age x , empirical and theoretical pointwise confidence intervals can be constructed.

If the true model is known, the pointwise mean and standard deviation of $g(x, \hat{\boldsymbol{\theta}})$ can be estimated through resampling. This provides a confidence interval for each value of x , and the width of the resulting confidence band gives us an indication of the stability of the estimated model, given the true model. If the true model is unknown, bootstrapping from the data sample substitutes resampling from the true model.

Asymptotic theoretical confidence intervals are found using the delta method,

$$g(x, \hat{\boldsymbol{\theta}}) \sim \mathcal{N}(g(x, \boldsymbol{\theta}), \nabla g(x, \boldsymbol{\theta}) J(\hat{\boldsymbol{\theta}})^{-1} \nabla g(x, \boldsymbol{\theta})^\top), \quad \text{as } N \rightarrow \infty,$$

where

$$J(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

is the observed information matrix. A 95% pointwise confidence interval for $g(x, \boldsymbol{\theta})$ is then given by

$$g(x, \hat{\boldsymbol{\theta}}) \pm 1.96 \sqrt{\nabla g(x, \hat{\boldsymbol{\theta}}) J(\hat{\boldsymbol{\theta}})^{-1} \nabla g(x, \hat{\boldsymbol{\theta}})^\top}. \quad (4.12)$$

5 Numerical applications

In this section, the two parameter estimation methods are applied to three types of illustrative examples. We first analyse artificial examples in which we simulate ATMMPPs, then we estimate their parameters based on the simulations, we make a goodness of fit analysis and compare the different criteria for choosing the optimal number of phases. We provide a summary of the results here and refer to Section 7.4 in the Supplementary Material for details. Next, we use real global female population data in different countries to estimate the model parameters. Finally, we fit an MBT using real individual demographic data on the black robin population, and we give a biological interpretation to our results.

We used the Matlab function `fmincon` to minimize the sum of weighted squared errors (3.1) in the global population data case, or to maximize the log-likelihood function (4.2) in the individual demographic data case, under the constraint of positive parameters. The algorithm requires an initial value (seed) for the parameters. A reasonable guess for the model parameters is

$$\gamma_i = \frac{n}{M+1} \quad \text{for } 1 \leq i \leq n-1,$$

and

$$\lambda_i = \frac{\sum_{x=0}^M \hat{b}_x}{M+1}, \quad \mu_i = \frac{\sum_{x=0}^M \hat{d}_x}{M+1}, \quad \text{for } 1 \leq i \leq n.$$

In order to minimize the risk to converge to local extrema, we started the algorithm with 25 different seeds obtained by adding random noise to the above values, and we chose the optimal solution among all.

5.1 Artificial examples

We considered three examples of ATMMPPs with $n = 3$ or $n = 4$ phases and we applied the two parameter estimation methods on data samples constructed by simulating trajectories of these models. As expected, the fits corresponding to the individual demographic data are much closer to the real model, and are associated to smaller confidence bands, than those corresponding to the global population data.

We observe that the MSE does not seem to be a satisfactory criterion to determine the optimal number of phases as the real value of n never minimizes the MSE on these examples. In all cases, the AIC provides the

correct answer most of the time, while the CV and MSIL show similar trends and slightly under-estimate the true value of n . The parameters K and M in the MSIL were chosen according to (4.10) and the criterion turned out not to be sensitive to this choice as the optimal value of n is the same for neighbouring values of K and M . All the details and figures can be found in Section 7.4 in the Supplementary Material.

5.2 Female families in different countries

In [4] the authors used real global population data on female human mortality and fertility rates corresponding to five-year ages-classes from different countries to fit MBTs with 22 phases; in these models, each phase corresponds exactly to one age-class. Here we use the weighted non-linear regression method described in Section 7.2 to estimate the parameters of new MBTs with 22 phases, and we compare the model age-specific mortality and fertility curves resulting from both approaches. Besides facilitating the comparison, considering the same number of phases allows us to start the optimisation algorithm with the most realistic initial parameter values given by the model values in [4]. We show the results for a supercritical country (Congo), an almost-critical country (USA) and a subcritical country (Japan) in Figure 1. We see that the new fits have substantially improved: the MSE is divided by a factor of 5.21 for Congo, 1.89 for the USA and 1.38 for Japan. We observe that the fits of the mortality curves are less satisfactory for the older ages; note that removing the weights (which are decreasing with age) do not improve the fits.

5.3 Black Robin population

The black robin is an endangered songbird species endemic to the Chatham Islands, an isolated archipelago located 800 kilometres East of New Zealand. By 1980, the population of black robins had declined to five birds, including only a single successful breeding pair, on Mangere Island [1]. Through intensive conservation efforts in 1980-1989 by the New Zealand Wildlife Service (now the Department of Conservation), the population recovered to 93 birds by spring 1990 [7]. Over the next decade (1990-1998), the population was closely monitored, but without human intervention. Nevertheless the population continued to grow rapidly to 197 adults by 1998, but after this

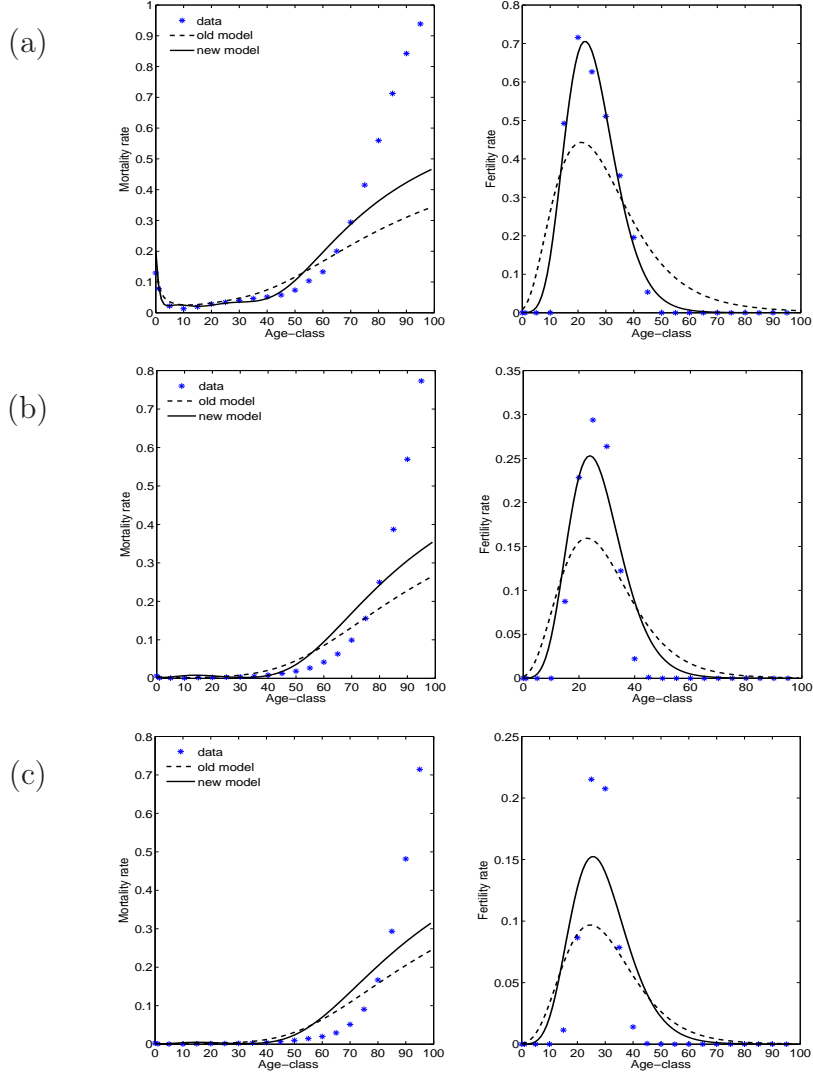


Figure 1: **Female families.** Optimal model with $n = 22$ computed using global population data (new model) corresponding to Congo (a), USA (b) and Japan (c), compared to the model used in [4] (old model).

period, the population growth slowed considerably and it only reached 239 adults in 2011 and 298 in 2014 [11].

For the conservation management of highly threatened species it is important to know the potential future viability, or survival probability, of a population, because if a population is not viable (i.e. fertility and survival

rates are low), it will eventually become extinct. Population viability depends on reproductive rates and survival of individuals, but these rates may vary between sexes and across an individual's life span (with age). Hence, the exact male-to-female ratio and the ages of each individual within a population will influence a population's future viability. Reintroduction of a species into previously occupied parts of its former natural range is nowadays a common hands-on conservation method. In order to maximize the survival chances of the new population, it is necessary to know the optimal age distribution of the reintroduced population, which can only be designed based on a complete age-specific demographic analysis of the species. The black robin is an ideal species for which to develop these new statistical tools, because biologists have been collecting complete raw datasets on this bird species for several decades. An age-specific reorganisation of these data leads to a total of 433 life vectors for the monitoring period 2007-2014.

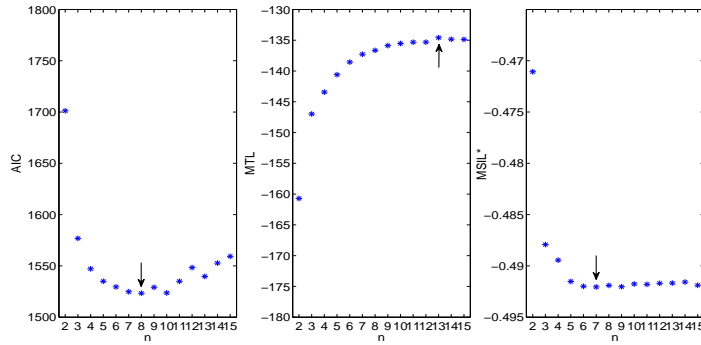


Figure 2: **Black robins.** Left: Result of AIC: the optimal number of phases is $n = 8$; Middle: Result of CV: the optimal number of phases is $n = 13$; Right: Result of MSIL with $M = 3$ and $K = 1$: the optimal number of phases is $n = 7$.

We performed different tests to determine the optimal number of phases to fit the black robin data. The results are shown in Figure 2, where we see that the optimal value is $n = 8$ according to the AIC, $n = 13$ according to the CV criterion, and $n = 7$ according to the MSIL criterion with $M = 3$ and $K = 1$ (determined using (4.10)). In this case, the MSIL criterion is quite sensitive to the choice of M and K , as indicated in Table 1.

The model fits based on the global population data and on the individual

$M \setminus K$	0	1	2	3
2	5	9	13	13
3	5	7	14	14
4	5	15	14	14

Table 1: **Black robins.** The optimal number of phases according to the MSIL criterion for different values of M and K .

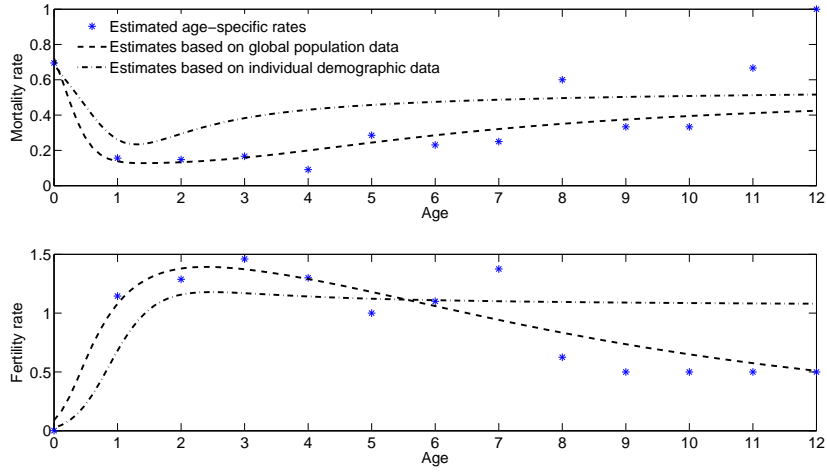


Figure 3: **Black robins.** Age-specific mortality and fertility curves for the models based on global population data and on individual demographic data with $n = 8$.

demographic data (life vectors) with $n = 8$ are compared in Figure 3. Since life vectors are available here, the corresponding models are the most informative. Black robins reach maturity at 1 or 2-years of age. The age-specific mortality curves show that mortality of black robins is the highest before they reach maturity and the lowest when they are 1 to 2 years old. Once birds reach 3 years of age, mortality rates do not increase dramatically with age, nor do fertility rates decline, which would support the hypothesis that there is no senescence. However, as few birds reach the old ages, the accuracy of the estimates obtained using global population data declines with age. Figure 4 shows the 95% pointwise confidence intervals for the estimates of the model outputs, obtained by bootstrapping 25 samples from the original

sample of life vectors. We see that the confidence intervals are quite narrow, especially for the estimates obtained using the individual demographic data.

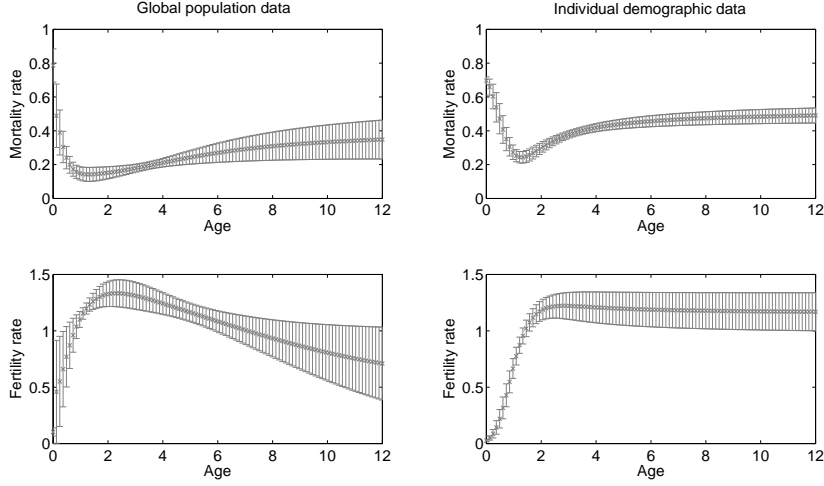


Figure 4: **Black robins.** Mean and 95% pointwise confidence intervals of the model fits corresponding to 25 bootstrapped datasets generated from the individual dataset containing $N = 433$ life vectors, for $n = 8$, using global population data (right) and individual demographic data (left).

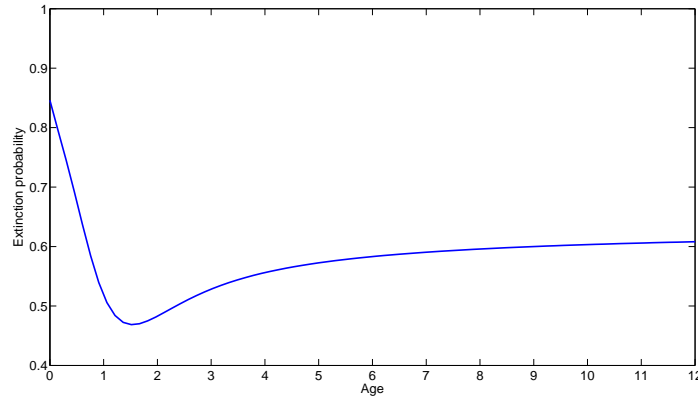


Figure 5: **Black robins.** The extinction probability of a female population as a function of the age of the initial female computed from the model with $n = 8$ estimated using individual demographic data.

One of the most informative model output is the probability of extinction of a female family (that is, consisting only of female descendants) generated by a single female, as a function of the age of that first female. This probability can be computed from the MBT model using any of the available algorithms (see for instance [5, 6]) and is shown in Figure 5. The curve highlights the combined effect of the age-specific mortality and fertility rates on the viability of the female family, hence of the whole population, by extension. We see that, if we were to found a new population starting with a single female bird, in order to maximize the survival chance of the population, the optimal age of the initial female should be around one year old.

6 Future directions

There are a number of directions for future research, particularly for the study of global population data. Cross-validation is delicate in that case as there are generally few data points. Leave-one-out cross-validation could be used, where we leave one age-class out. Possible complications would then include the choice of the weights \hat{S}_x as these explicitly use the death rates for all age-classes.

Further methods for analysing global population data could be developed for when the age-specific mortality and fertility rates follow some specific distributions. For example, we may know that the birth rate and death rate at each age lie in an exponential family of distributions. The analysis could then involve a parallel process of estimating these distributions and using these distributions to simulate new samples of global population data. Each new sample can then be used to do parameter estimation and construct confidence intervals. Alternatively, the theory developed for finding confidence intervals for weighted non-linear regression methods could be used in this case; unfortunately, this process is not computationally straightforward.

Acknowledgements

Sophie Hautphenne thanks the Australian Research Council (ARC) for support through the Discovery Early Career Researcher Award DE150101044. The black robin research was funded by the New Zealand Foundation for Research, Science and Technology (UOCX0601) to Melanie Massaro, by the School of Biological

Sciences, University of Canterbury, by the Brian Mason Scientific and Technical Trust, and by the Mohamed bin Zayed Species Conservation Fund. This research was only possible with permission from the Chatham Island Conservation Board and the logistic help of the Department of Conservation.

References

- [1] D. Butler and D. Merton. *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press, Auckland, 1992.
- [2] A. C. Davison and N. I. Ramesh. Some models for discretized series of events. *Journal of the American Statistical Association*, 91(434): 601–609, 1996.
- [3] T. B. Gage and C. J. Mode. Some laws of mortality: how well do they fit? *Human Biology*, 65(3) 445–461, 1993.
- [4] S. Hautphenne and G. Latouche. The Markovian binary tree applied to demography. *Journal of mathematical biology*, 64(7):1109–1135, 2012.
- [5] S. Hautphenne, G. Latouche, and M.-A. Remiche. Newton’s iteration for the extinction probability of a Markovian Binary Tree. *Linear Algebra and its Applications*, 428:2791–2804, 2008.
- [6] S. Hautphenne, G. Latouche, and M.-A. Remiche. Algorithmic approach to the extinction probability of branching processes. *Methodology and Computing in Applied Probability*. 13(1), 171–192, 2011.
- [7] E. S. Kennedy, C. E. Grueber, R. P. Duncan, and I. G. Jamieson. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species—the Chatham Island black robin. *Evolution*, 68(4):987–995, 2014.
- [8] G. Latouche, M.-A. Remiche, and P. Taylor. Transient Markov arrival processes. *Annals of Applied Probability*, pages 628–640, 2003.
- [9] X.S. Lin and X. Liu. Markov aging process and phase-type law of mortality. *North American Actuarial Journal*, 11:92–109, 2007.
- [10] M. Massaro, R. Sainudiin, D. Merton, J.V. Briskie, A.M. Poole, and M.L. Hale. Human-assisted spread of a maladaptive behavior in a critically endangered bird. *PloS one*, 8(12), e79066, 2013.
- [11] M. Massaro, M. Stanbury, and J. V. Briskie. Nest site selection by the endangered black robin increases vulnerability to predation by an invasive bird. *Animal Conservation*, 16(4):404–411, 2013.

- [12] D Merton. The chatham island black robin. *Forest and bird*, 21(3):14–19, 1990.
- [13] P. Peristera and A. Kostaki. Modeling fertility in modern populations. *Demographic Research*, 16(4):141–194, 2007.
- [14] N. I. Ramesh. Statistical analysis on Markov-modulated poisson processes. *Environmetrics*, 6(2):165–179, 1995.
- [15] T. Rydén. Parameter estimation for Markov modulated Poisson processes. *Stochastic Models*, 10(4):795–829, 1994.

7 Supplementary Material

7.1 Proof of Proposition 3.1

Let L be the lifetime of an individual and $\bar{S}(x) = P[L > x]$ be the survival function. Since L follows a $\text{PH}(\boldsymbol{\alpha}, D)$ distribution,

$$\bar{S}(x) = 1 - P[L \leq x] = \boldsymbol{\alpha} e^{Dx} \mathbf{1}.$$

The probability of death at age x , $\bar{d}(x)$, can thus be calculated as

$$\begin{aligned} \bar{d}(x) &= P[x < L \leq x+1 | L > x] = \frac{P[L > x] - P[L > x+1]}{P[L > x]} \\ &= \frac{\bar{S}(x) - \bar{S}(x+1)}{\bar{S}(x)} = \frac{\boldsymbol{\alpha} e^{Dx} (I - e^D) \mathbf{1}}{\boldsymbol{\alpha} e^{Dx} \mathbf{1}}. \end{aligned}$$

It is shown in [8] that the mean number of events until time t in a TMAP started in phase i at time 0 is given by

$$E[N(t) | \varphi(0) = i] = [(I - e^{Dt})(-D)^{-1} D_1 \mathbf{1}]_i.$$

Let $N([x, x+t])$ denote the number of events in the TMAP in the time interval $[x, x+t]$. By time-homogeneity of the TMAP,

$$E[N([x, x+t]) | \varphi(x) = i] = E[N(t) | \varphi(0) = i].$$

The mean number of offspring generated by an individual at age x can thus be calculated as

$$\begin{aligned} \bar{b}(x) &= E[N([x, x+1]) | L > x] \\ &= \sum_{1 \leq i \leq n} P[\varphi(x) = i | L > x] E[N([x, x+1]) | L > x, \varphi(x) = i] \\ &= \sum_{1 \leq i \leq n} \frac{P[\varphi(x) = i, L > x]}{P[L > x]} E[N(1) | \varphi(0) = i] \\ &= \sum_{1 \leq i \leq n} \frac{[\boldsymbol{\alpha} e^{Dx}]_i}{\boldsymbol{\alpha} e^{Dx} \mathbf{1}} [(I - e^D)(-D)^{-1} D_1 \mathbf{1}]_i \\ &= \frac{\boldsymbol{\alpha} e^{Dx} (I - e^D)(-D)^{-1} D_1 \mathbf{1}}{\boldsymbol{\alpha} e^{Dx} \mathbf{1}}. \end{aligned}$$

□

7.2 Global population data with ℓ -year age classes

In demography, the available data often consist of age-specific fertility and mortality rates over ℓ -year age-classes with $\ell > 1$, that is,

- the expected number of offspring *per year* from a parent in age-class $[x, x + \ell)$, denoted as $\hat{\beta}_{[x, x + \ell)}$, and
- the probability that an individual who reached the age-class $[x, x + \ell)$ dies *within the year*, denoted as $\hat{\mu}_{[x, x + \ell)}$.

We extend the definition of $\bar{d}(x)$ and $\bar{b}(x)$ to ℓ -year age-classes and define the functions $\bar{d}(x, \ell)$ and $\bar{b}(x, \ell)$, computed from the TMAP model, as follows:

$$\begin{aligned}\bar{d}(x, \ell) &= P[x < L \leq x + \ell | L > x] \\ \bar{b}(x, \ell) &= E[N([x, x + \ell)) | L > x].\end{aligned}$$

It is a simple matter to generalise Proposition 3.1 to obtain

Corollary 7.1.

$$\begin{aligned}\bar{d}(x, \ell) &= \frac{\alpha e^{Dx} (I - e^{D\ell}) \mathbf{1}}{\alpha e^{Dx} \mathbf{1}} \\ \bar{b}(x, \ell) &= \frac{\alpha e^{Dx} (I - e^{D\ell}) (-D)^{-1} D \mathbf{1}}{\alpha e^{Dx} \mathbf{1}}.\end{aligned}$$

The correspondence between $\bar{d}(x, \ell)$ and $\bar{b}(x, \ell)$ and the mortality and fertility rates in age-class $[x, x + \ell)$ is given in the next Lemma.

Lemma 7.1.

$$\begin{aligned}\bar{d}(x, \ell) &\equiv 1 - (1 - \hat{\mu}_{[x, x + \ell)})^\ell \\ \bar{b}(x, \ell) &\equiv \hat{\beta}_{[x, x + \ell)} \frac{1 - (1 - \hat{\mu}_{[x, x + \ell)})^\ell}{\hat{\mu}_{[x, x + \ell)}},\end{aligned}$$

where the symbol \equiv has to be interpreted as “is the model equivalent of”.

Proof. The model function $1 - \bar{d}(x, \ell)$ is the probability that an individual who reached age x survives at least until age $x + \ell$, that is, survives ℓ successive one-year age intervals, which occurs with probability $(1 - \hat{\mu}_{[x, x + \ell)})^\ell$.

The model function $\bar{b}(x, \ell)$ can be rewritten as

$$\begin{aligned}\bar{b}(x, \ell) &= E[N([x, x + 1)) | L > x] + E[N([x + 1, x + 2)) | L > x] + \dots \\ &\quad + E[N([x + \ell - 1, x + \ell)) | L > x] \\ &\equiv \hat{\beta}_{[x, x + \ell)} + (1 - \hat{\mu}_{[x, x + \ell)}) \hat{\beta}_{[x, x + \ell)} + \dots + (1 - \hat{\mu}_{[x, x + \ell)})^{\ell-1} \hat{\beta}_{[x, x + \ell)} \\ &= \hat{\beta}_{[x, x + \ell)} \frac{1 - (1 - \hat{\mu}_{[x, x + \ell)})^\ell}{\hat{\mu}_{[x, x + \ell)}},\end{aligned}$$

which completes the proof. □

In order to estimate the model parameters in this case, the objective function (3.1) then needs to be modified according to Corollary 7.1 and Lemma 7.1.

7.3 Proof of Proposition 4.1

In order to compute $P(k)$, we actually compute $P(k, t)$ for $0 \leq k \leq K$, and for any $t \geq 0$ where

$$P_{ij}(k, t) := P[N(t) = k, \varphi(t) = j | N(0) = 0, \varphi(0) = i],$$

and observe that $P(k) = P(k, \ell)$. It is well known from the theory of MAPs that the probability generating function $P^*(z, t) := \sum_{k \geq 0} P(k, t) z^k$, is given by the matrix exponential

$$P^*(z, t) = \exp[D(z)t], \quad \text{where } D(z) := D_0 + z D_1.$$

Since $P(k, t) = (1/k!) [\partial^k P^*(z, t) / (\partial z)^k] \big|_{z=0}$ for any $k \geq 0$, we need to take the derivatives of the matrix exponential $\exp[D(z)t]$ with respect to z . The scalar rule of exponential differentiation only holds here if D_0 and D_1 commute, which is generally not the case. Instead, we first differentiate $P^*(z, t)$ with respect to t ,

$$\partial P^*(z, t) / \partial t = D(z) P^*(z, t), \tag{7.1}$$

and we then take successive derivatives of this equation with respect to z :

$$\begin{aligned} \partial^2 P^*(z, t) / (\partial t)(\partial z) &= D_1 P^*(z, t) + D(z) \partial P^*(z, t) / \partial z \\ \partial^3 P^*(z, t) / (\partial t)(\partial z)^2 &= 2D_1 \partial P^*(z, t) / \partial z + D(z) \partial^2 P^*(z, t) / (\partial z)^2 \\ &\vdots \\ \partial^{(K+1)} P^*(z, t) / (\partial t)(\partial z)^K &= K D_1 \partial^{(K-1)} P^*(z, t) / (\partial z)^{(K-1)} \\ &\quad + D(z) \partial^K P^*(z, t) / (\partial z)^K. \end{aligned}$$

This system of partial derivative equations can be rewritten as an ordinary differential equation for the unknown matrix containing the partial derivatives of $P^*(z, t)$ with respect to z ,

$$\frac{d}{dt} \begin{bmatrix} P^*(z, t) \\ \partial P^*(z, t) / \partial z \\ \partial^2 P^*(z, t) / (\partial z)^2 \\ \vdots \\ \partial^K P^*(z, t) / (\partial z)^K \end{bmatrix} = \begin{bmatrix} D(z) & & & & \\ D_1 & D(z) & & & \\ & 2D_1 & D(z) & & \\ & & \ddots & \ddots & \\ & & & K D_1 & D(z) \end{bmatrix} \begin{bmatrix} P^*(z, t) \\ \partial P^*(z, t) / \partial z \\ \partial^2 P^*(z, t) / (\partial z)^2 \\ \vdots \\ \partial^K P^*(z, t) / (\partial z)^K \end{bmatrix},$$

whose solution is

$$\begin{bmatrix} P^*(z, t) \\ \partial P^*(z, t)/\partial z \\ \partial^2 P^*(z, t)/(\partial z)^2 \\ \vdots \\ \partial^K P^*(z, t)/(\partial z)^K \end{bmatrix} = \exp \left(\begin{bmatrix} D(z) & & & \\ D_1 & D(z) & & \\ & 2D_1 & D(z) & \\ & & \ddots & \\ & & & KD_1 & D(z) \end{bmatrix} t \right) \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Taking $z = 0$ and denoting

$$\mathcal{M} = \begin{bmatrix} D_0 & & & \\ D_1 & D_0 & & \\ & 2D_1 & D_0 & \\ & & \ddots & \\ & & & KD_1 & D_0 \end{bmatrix},$$

we obtain

$$P(k) = P(k, \ell) = (1/k!)(\mathbf{e}_k \otimes I) \exp(\mathcal{M}\ell)(\mathbf{e}_1^\top \otimes I).$$

Then, $\mathbf{p}(k)$ is obtained by conditioning on the time $u \in [0, \ell]$ when the individual dies,

$$\begin{aligned} \mathbf{p}(k) &= \int_0^\ell P(k, u) \mathbf{d} du \\ &= (1/k!)(\mathbf{e}_k \otimes I) \int_0^\ell \exp(\mathcal{M}u) du (\mathbf{e}_1^\top \otimes I) \mathbf{d} \\ &= (1/k!)(\mathbf{e}_k \otimes I) [I - \exp(\mathcal{M}\ell)](-\mathcal{M})^{-1} (\mathbf{e}_1^\top \otimes I) \mathbf{d} \end{aligned}$$

Next,

$$P = P^*(1, \ell) = \exp(D\ell),$$

where $D = D_0 + D_1$, and finally

$$\mathbf{p} = \int_0^\ell \exp(Du) \mathbf{d} du = [I - \exp(D\ell)](-D)^{-1} \mathbf{d}.$$

□

7.4 Further details on the artificial examples

We consider three examples of ATMMPPs, simulating N trajectories of these processes for T units of time. The different parameter values are summarized in Table 2.

	n	γ_1	γ_2	γ_3	μ_1	μ_2	μ_3	μ_4	λ_1	λ_2	λ_3	λ_4	N	T
Ex. 1	3	0.25	0.25	—	0.2	0.4	0.9	—	6	3	2	—	500	15
Ex. 2	4	0.5	0.1	0.1	0.3	0.1	0.2	0.7	0.5	2	0.5	0.01	400	25
Ex. 3	4	0.3	0.3	0.3	0.6	0.1	0.2	0.5	0.2	3	2	0.1	500	15

Table 2: Parameter values and simulation characteristics of the ATMMPPs corresponding to the three artificial examples. The number of phases is denoted by n , γ_i is the transition rate from phase i to phase $i + 1$, μ_i is the death rate in phases i , λ_i is the birth rate in phase i , N is the number of simulated trajectories, and T is the simulation time for each trajectory.

For each example, we associated each simulated trajectory of the ATMMPP with a life vector by counting the number of births falling in successive ℓ -year intervals. Here we took $\ell = 1$, so each entry of the vectors corresponds to a specific age-classes and the vectors have a total of T entries. This produced samples $\{\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(N)}\}$ of N individual life vectors of the form (4.1). The average age-specific fertility and mortality rates \hat{b}_x and \hat{d}_x , for $0 \leq x \leq M = T - 1$, were computed directly from these samples.

We first performed a goodness of fit analysis on Example 1. We set $n = 3$ (the true number of phases), and we used (i) the global population data \hat{d}_x and \hat{b}_x , and (ii) the full sample of N life vectors, to estimate the model parameters using the corresponding statistical method, leading to two different parameter estimates $\hat{\theta}^{(i)}$ and $\hat{\theta}^{(ii)}$. In Figure 6 we compare the performance measures $g(x, \hat{\theta}^{(i)})$ and $g(x, \hat{\theta}^{(ii)})$ corresponding to the age-specific mortality and fertility curves, obtained with the two different estimation methods. To further assess the accuracy of the estimates, we re-sampled 50 datasets of size N from the true model, and we show in Figure 7 the mean curves compared to the real ones, as well as the corresponding 95% pointwise confidence intervals. In Figure 8, we perform the same analysis by bootstrapping 50 times from a single dataset instead of resampling. We conclude from Figures 6, 7 and 8 that, as expected, the fits corresponding to the individual demographic data are much closer to the real model, and are associated to smaller confidence bands, than those corresponding to the global population data. In Figure 9 we show the theoretical 95% pointwise confidence intervals given by (4.12) for the fits obtained using individual demographic data; these are comparable to those shown in Figure 7. Finally, in Figure 10 we compare the fits based on life vectors with different age-class lengths ℓ to those based on the observation of the successive inter-event times. We see that as ℓ decreases to zero, the estimates obtained with our method converge to those based on the successive inter-event times.

For all three examples, Figure 11 shows the sensitivity of the MSE with respect to the number of phases in the fitted model. This also highlights the fact that the MSE does not seem to be a satisfactory criterion to determine the optimal number of phases as the real value of n never minimizes the MSE on these examples. Finally, Figure 12 compares all criteria to decide upon the optimal number of phases in the individual demographic data case. In all cases, the AIC provides the correct answer most of the time, while the CV and MSIL show similar trends and slightly underestimate the true value of n . The parameters K and M in the MSIL were chosen according to (4.10) and the criterion turned out not to be sensitive to this choice as the optimal value of n is the same for neighbouring values of K and M .

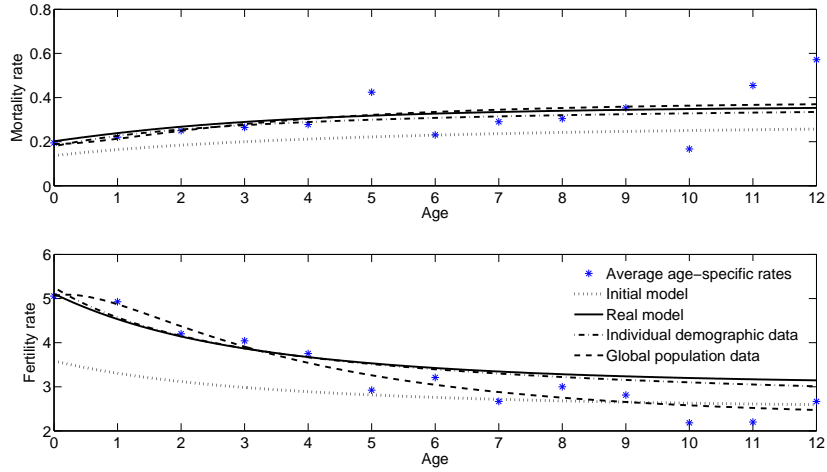


Figure 6: **Example 1.** Comparison of the model fits obtained using global population data and individual demographic data. The initial model is the one used as a seed in the optimisation algorithms.

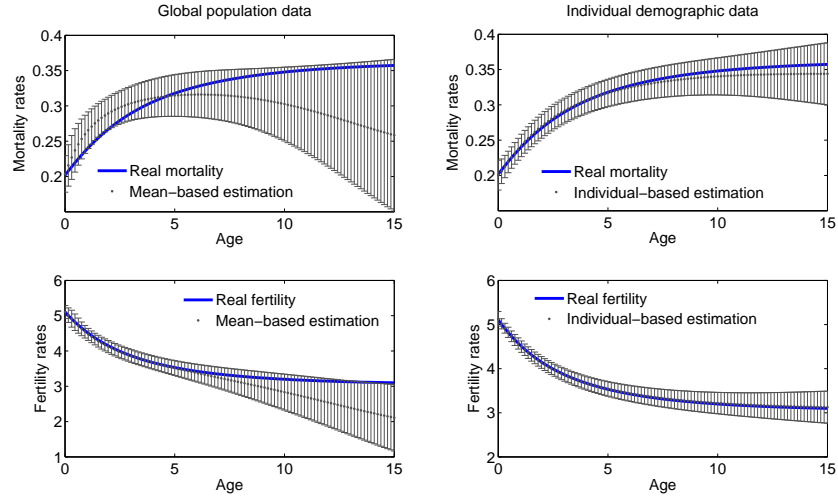


Figure 7: **Example 1.** Mean and 95% pointwise confidence intervals of the model fits corresponding to 50 simulations from the real model using global population data (left) and individual demographic data (right).

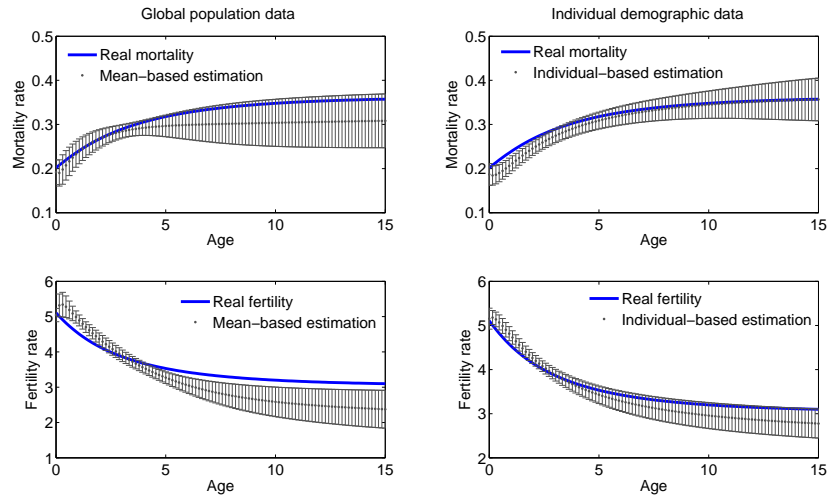


Figure 8: **Example 1.** Mean and 95% pointwise confidence intervals of the model fits corresponding to 50 bootstrapped datasets using global population data (left) and individual demographic data (right).

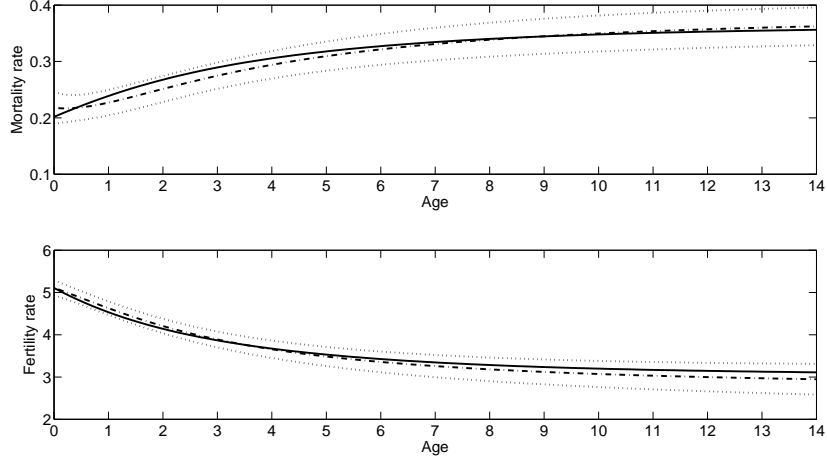


Figure 9: **Example 1.** Theoretical 95% pointwise confidence intervals (dotted lines) in the individual demographic data case. The true curves correspond to the plain lines and the estimated curves correspond to the dash-dot lines.

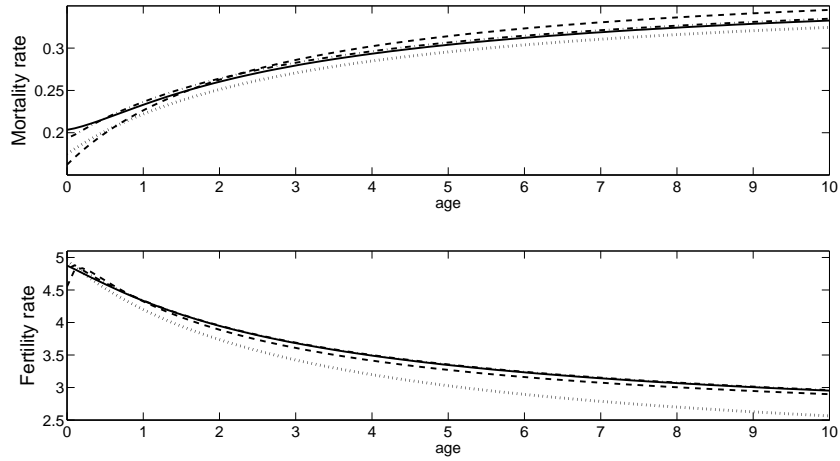


Figure 10: **Example 1.** Comparison of the model fits obtained using the MLE based on inter-arrival times (plain lines), and individual demographic data with $\ell = 5$ (dotted lines), $\ell = 2.5$ (dashed lines), and $\ell = 1$ (dash-dot lines).

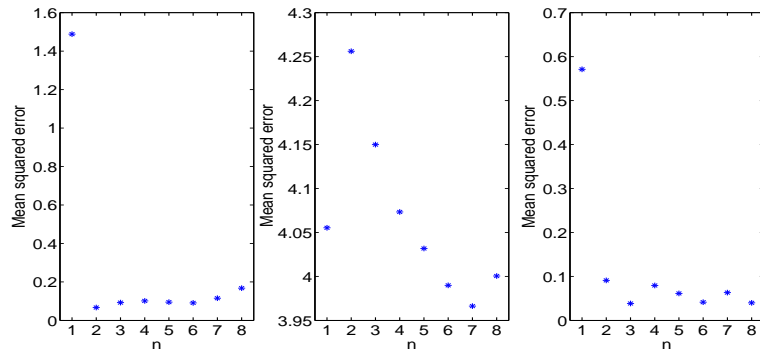


Figure 11: **Examples 1 (left), 2 (middle), and 3 (right).** Mean squared error based on 50 simulations from the true models.

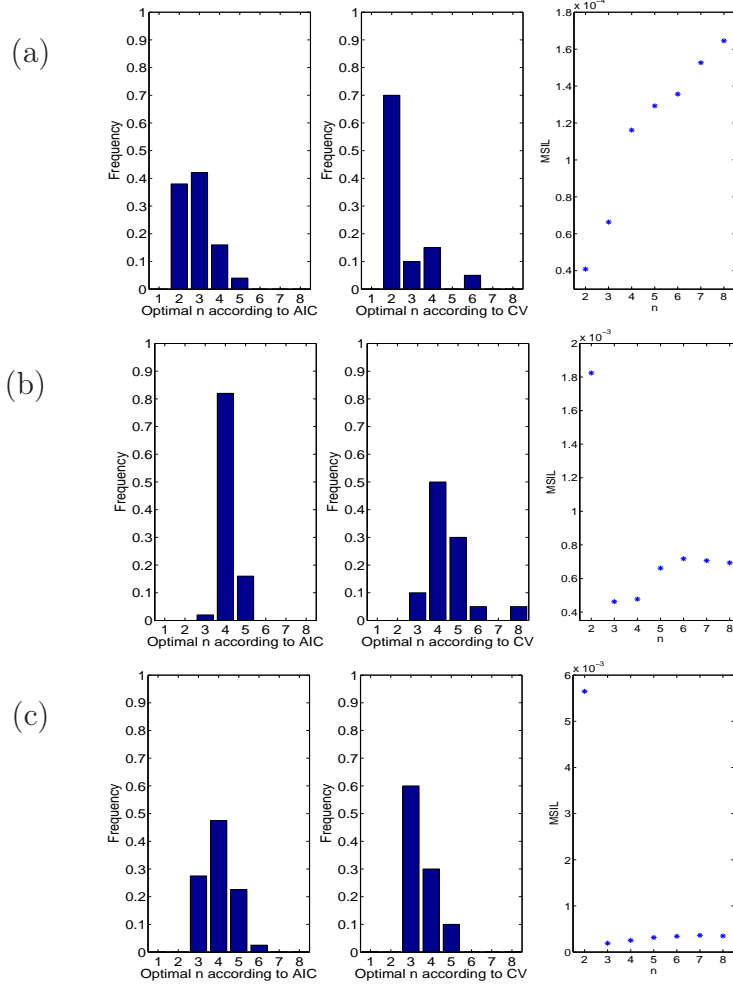


Figure 12: **Examples 1 (a), 2 (b), and 3 (c).** Left: Frequency of optimal n according to AIC based on 50 simulations from the true model. Middle: Frequency of optimal n according to CV based on 20 simulations. Right: MSIL for $2 \leq n \leq 8$ based on 50 simulations from the true model, with $M = 4$ and $K = 5$ (a), $M = 3$ and $K = 1$ (b), and $M = 3$ and $K = 2$ (c). We omit the value at $n = 1$ which is much larger than the value at $n = 2$.